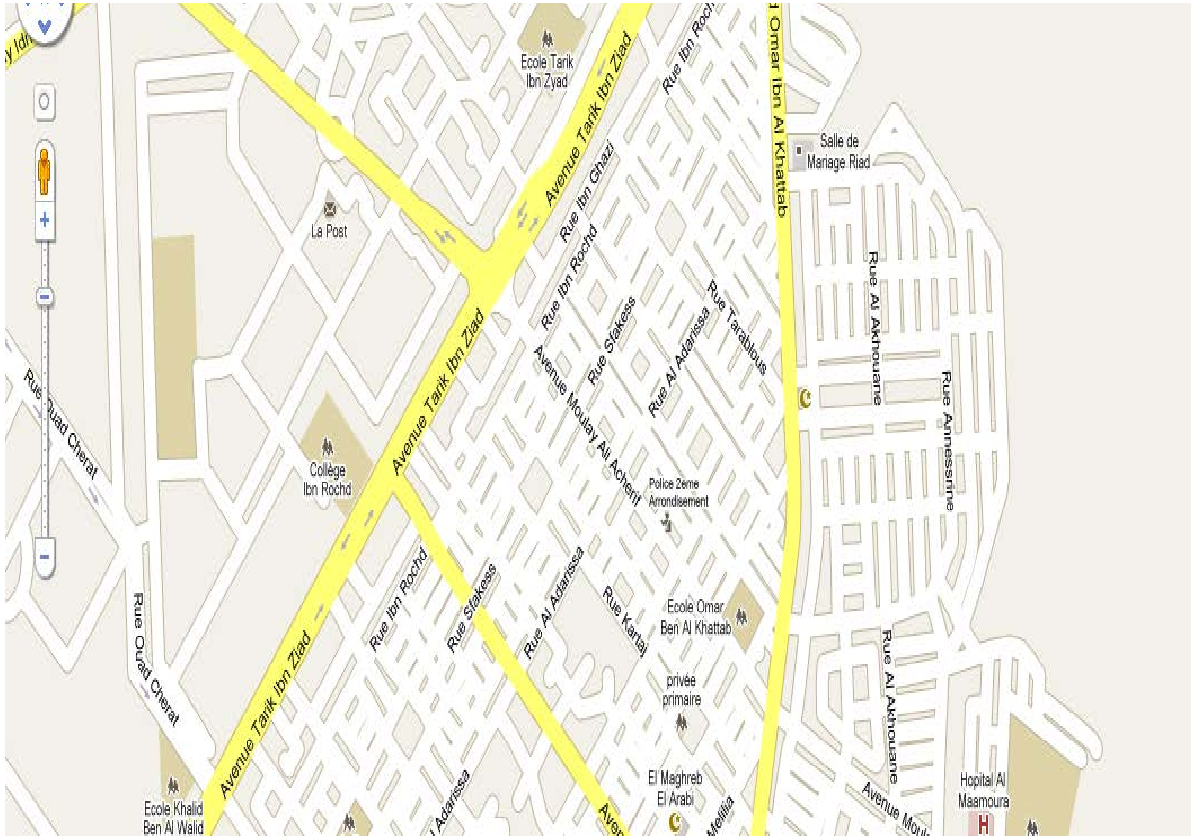
Gaussian Process based Radio Map Recovery

HuangZili

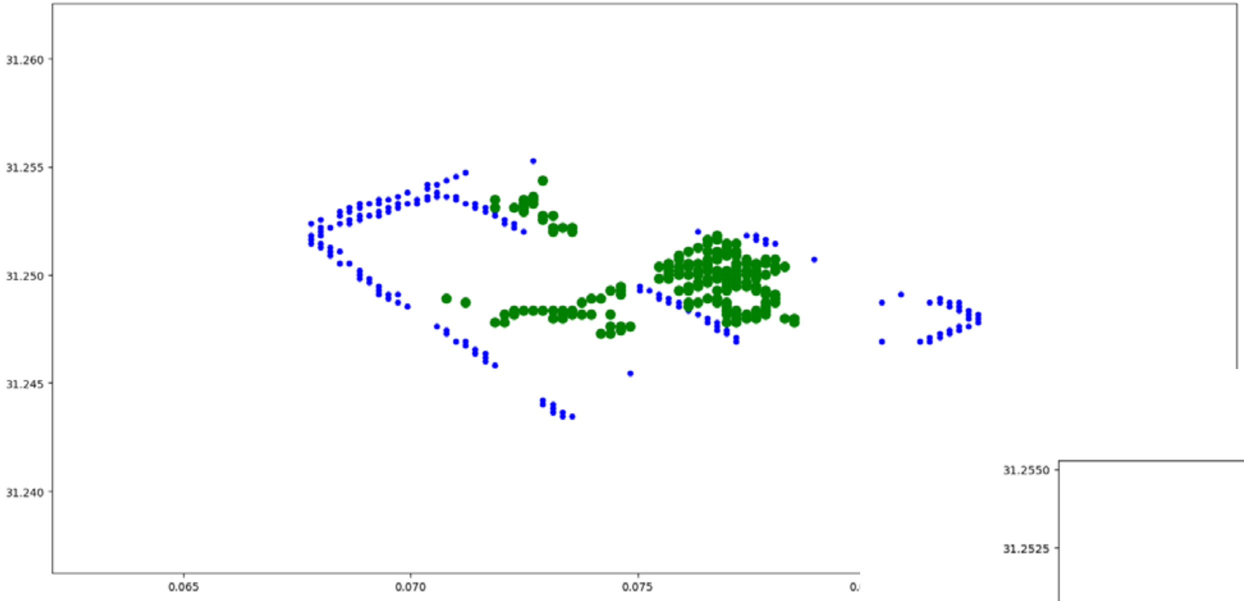
Content

Background

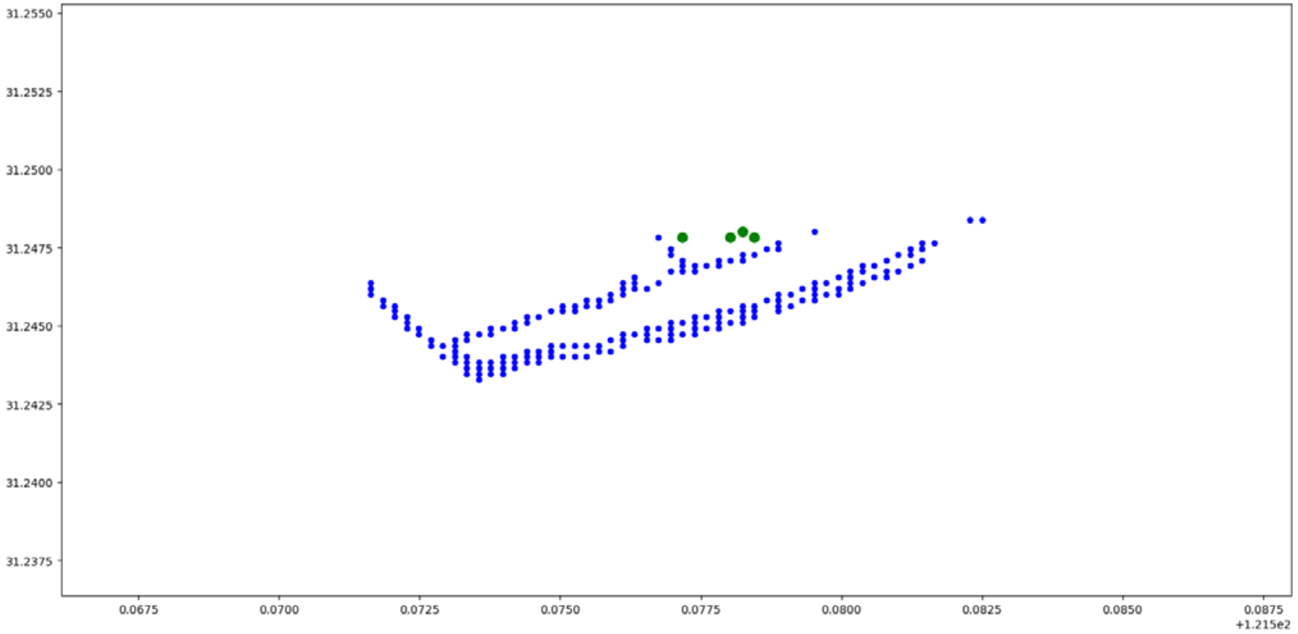
- To model the signal strength in one area, we can collect data from the users' mobile devices.
- The users are not uniformly distributed on the map. Some places like small streets may not be covered.
- How can we solve this problem?
- To predict the signal strength on the small roads based on signal strength on the main roads



Background



The test set is not generated by randomly sampling in the feature space.



Blue dots:
training samples
Green dots:
test samples

Method

Gaussian Process Regression

joint distribution

$$\begin{bmatrix} \mathbf{y} \\ f(\mathbf{x}_*) \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K}(\mathbf{X},\mathbf{X}) + \sigma_n^2\mathbf{I} & \mathbf{k}(\mathbf{X},\mathbf{x}_*) \\ \mathbf{k}(\mathbf{x}_*,\mathbf{X}) & k(\mathbf{x}_*,\mathbf{x}_*) \end{bmatrix}\right)$$

marginal distribution

$$f(\mathbf{x}_*) = \mathbf{k}_*^T \left(\mathbf{K} + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{y} = \mathbf{k}_*^T\boldsymbol{\alpha}\,,$$

$$V(\mathbf{x}_*) = k(\mathbf{x}_*,\mathbf{x}_*) - \mathbf{k}_*^T\left(\mathbf{K} + \sigma_n^2\mathbf{I}\right)^{-1}\mathbf{k}_*\,,$$

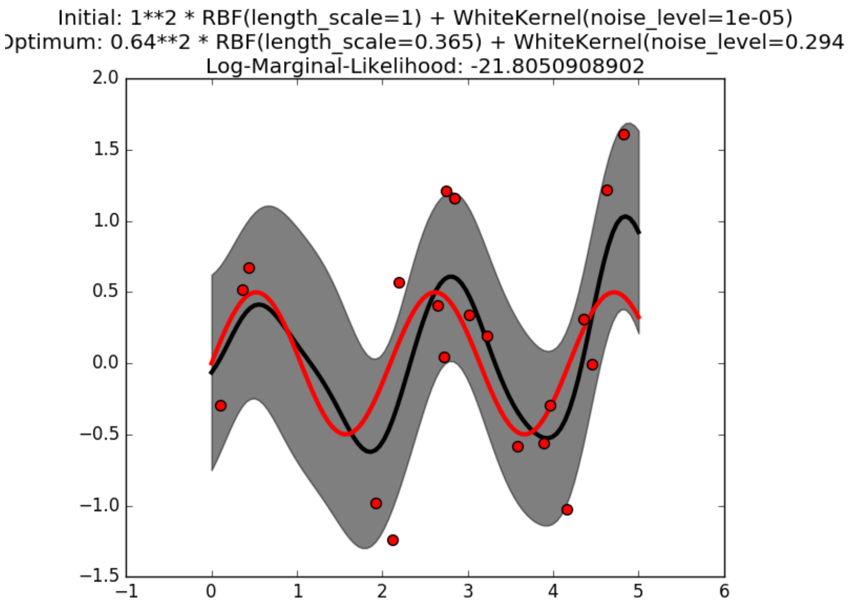
Method

Advantage of GPR

- (1)nonparametric
- (2)The prediction is probabilistic

Disadvantage of GPR

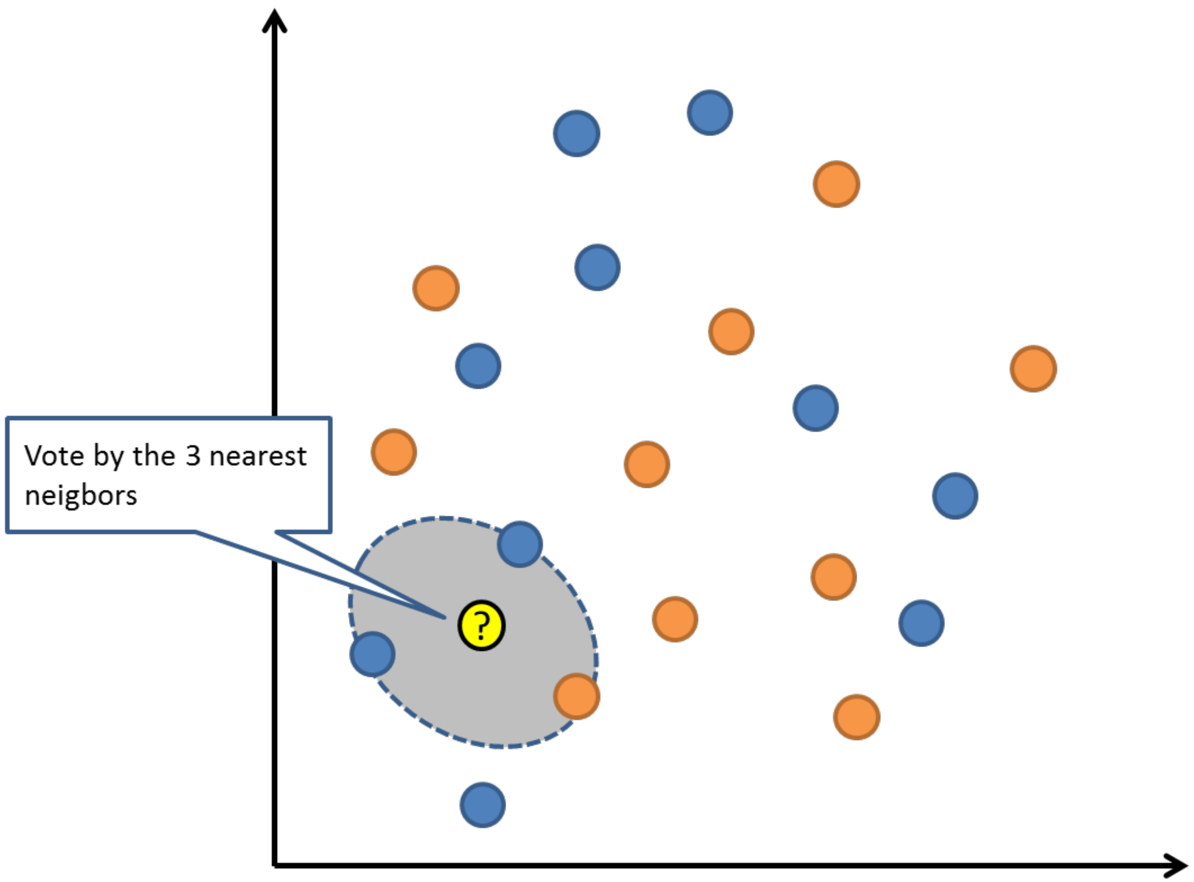
- (1)Gaussian processes are not sparse
- (2)They lose efficiency in high dimensional spaces



Method

K-nearest neighbors

- 1.distance measure
- 2.combination methods



Method

✚KNN-GPR

✚ Assumption

The signal strength follows Gaussian distribution in a local area.

✚ Process

- 1.Preprocess the data
- 2.Find k-nearest neighbors of the test sample
- 3.Apply Gaussian process regression in the k-nearest neighbors

Performance on our dataset

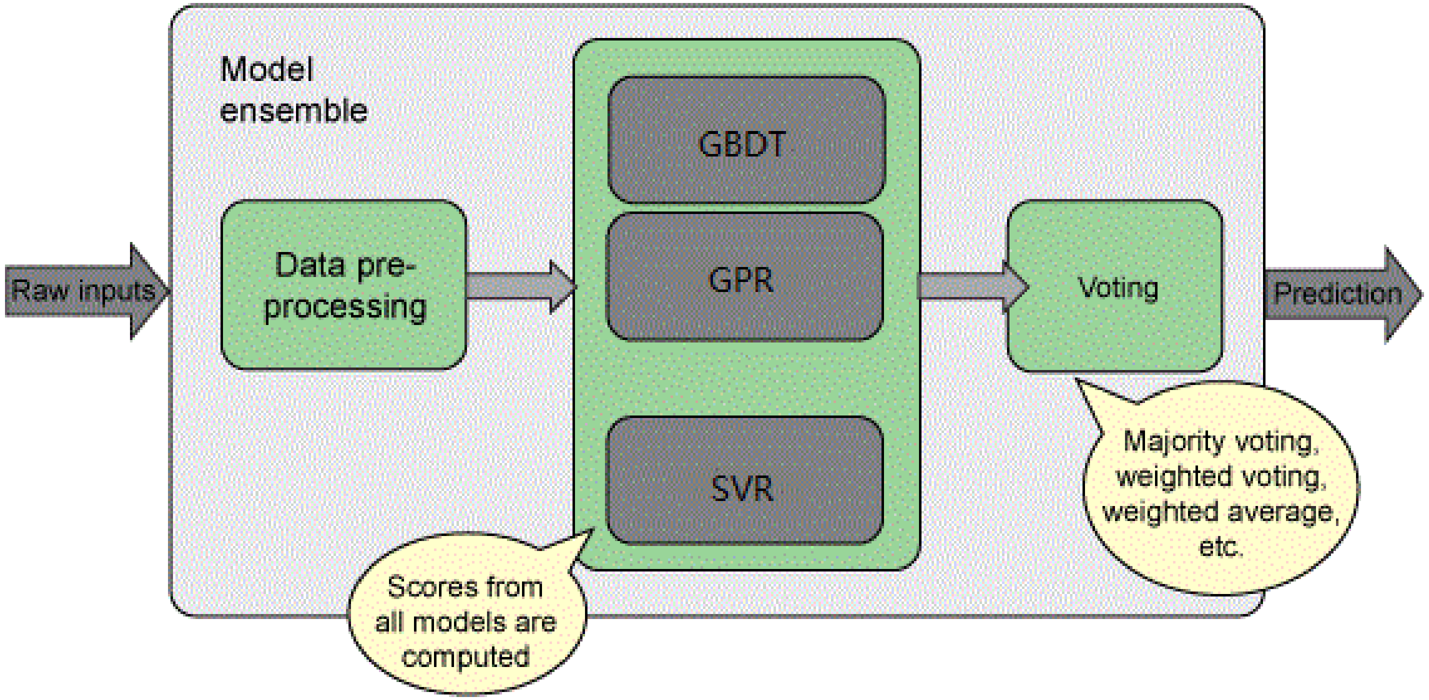
Methods	KNN(k=5)	KNN(k=10)	KNN(k=15)	KNN(k=30)	GP	KNN-GP
MAE	6.56	6.78	6.91	7.45	6.28	5.87

comparison of different methods

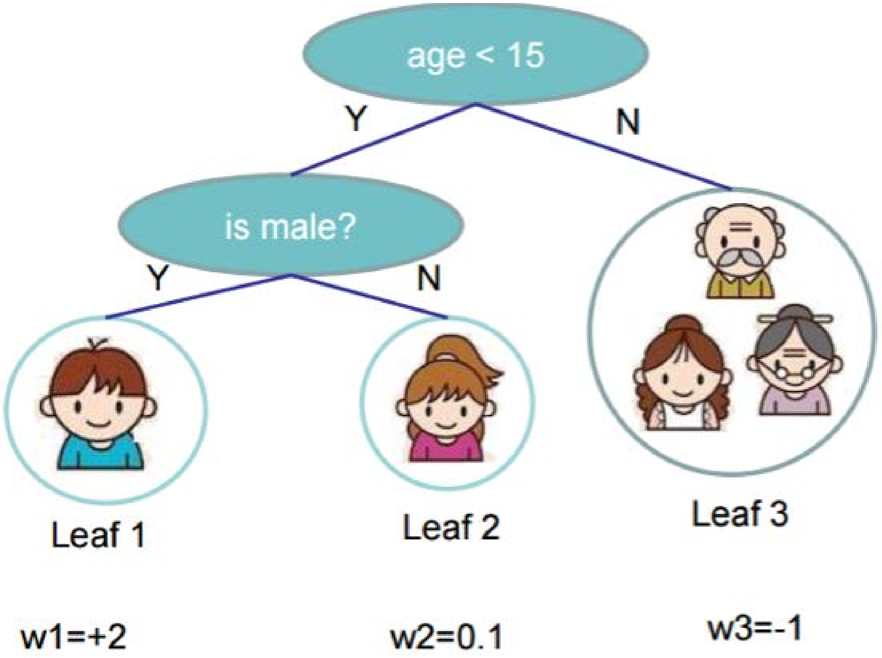
kernel	RBF	Matern(nu=0.5)	Matern(nu=1.5)	Matern(nu=2.5)	RationalQuadratic
MAE	5.87	6.98	6.15	6.04	7.84

comparison of different kernels

Method



GBDT: Gradient Boosted Decision Trees



Method

Methods	Gaussian Process Regression	Gradient Boosted Decision Trees	Support Vector Regression	Model ensemble(linear regression)	Model ensemble(average)
MAE	8.20	8.38	10.51	8.88	8.02

System

- A Gaussian Process Positioning System

$$\hat{t} = \arg\max_{t} \prod_{j:s_j \neq \{\}} p_j(s_j \mid t)$$

- $p_j(s_j \mid t)$ the likelihood of receiving a signal strength s_j from the j-th base station on position t.

- $p_j(s_j \mid t)$ is given by the Gaussian process regression

System

scale(m)	50	100	150	200
accuracy	0.3	0.525	0.375	0.65

results of position prediction with
no less than 3 data records

number of records	≥3	≥4	≥5	≥6
accuracy	0.525	0.655	0.75	1

results of position prediction
with different number of data
records

THANKS